

EpiBuffer: Modeling Spatial Heterogeneity in Exposure Buffers and Risk

SpatialBuffers Statistical Model

$$Y_i(\mathbf{s}_j) | \mu_i(\mathbf{s}_j), \zeta \stackrel{\text{ind}}{\sim} f(y | \mu_i(\mathbf{s}_j), \zeta), \quad j = 1, \dots, m, \quad i = 1, \dots, n_j;$$

$$g\{\mu_i(\mathbf{s}_j)\} = \mathbf{O}_i(\mathbf{s}_j) + \mathbf{x}_i(\mathbf{s}_j)^T \boldsymbol{\beta} + z\{\mathbf{s}_j; \delta(\mathbf{s}_j)\} \theta(\mathbf{s}_j).$$

Likelihood Options

- Binomial likelihood with logit link function (likelihood_indicator = 0):

$$Y_i(\mathbf{s}_j) | p_i(\mathbf{s}_j) \stackrel{\text{ind}}{\sim} \text{Binomial}(\tilde{n}_i(\mathbf{s}_j), p_i(\mathbf{s}_j)), \quad \text{logit}\{p_i(\mathbf{s}_j)\} = \mathbf{x}_i(\mathbf{s}_j)^T \boldsymbol{\beta} + z\{\mathbf{s}_j; \delta(\mathbf{s}_j)\} \theta(\mathbf{s}_j);$$

- Gaussian likelihood with identity link function (likelihood_indicator = 1):

$$Y_i(\mathbf{s}_j) = \mathbf{x}_i(\mathbf{s}_j)^T \boldsymbol{\beta} + z\{\mathbf{s}_j; \delta(\mathbf{s}_j)\} \theta(\mathbf{s}_j) + \epsilon_i, \quad \epsilon_i | \sigma_\epsilon^2 \stackrel{\text{iid}}{\sim} \text{N}(0, \sigma_\epsilon^2);$$

- Negative binomial likelihood with logit link function (likelihood_indicator = 2):

$$Y_i(\mathbf{s}_j) | r, p_i(\mathbf{s}_j) \stackrel{\text{ind}}{\sim} \text{Negative Binomial}(r, p_i(\mathbf{s}_j)), \quad \text{logit}\{p_i(\mathbf{s}_j)\} = \mathbf{O}_i(\mathbf{s}_j) + \mathbf{x}_i(\mathbf{s}_j)^T \boldsymbol{\beta} + z\{\mathbf{s}_j; \delta(\mathbf{s}_j)\} \theta(\mathbf{s}_j).$$

Exposure Definitions

- Counts (exposure_definition_indicator = 0):

$$z\{\mathbf{s}_j; \delta(\mathbf{s}_j)\} = \sum_{k=1}^h 1(d_{jk} \leq \delta(\mathbf{s}_j));$$

- Spherical (exposure_definition_indicator = 1):

$$z\{\mathbf{s}_j; \delta(\mathbf{s}_j)\} = \sum_{k=1}^h 1(d_{jk} \leq \delta(\mathbf{s}_j)) \left(1 - 1.5 \left\{ \frac{d_{jk}}{\delta(\mathbf{s}_j)} \right\} + 0.5 \left\{ \frac{d_{jk}}{\delta(\mathbf{s}_j)} \right\}^3 \right);$$

- Presence/absence (exposure_definition_indicator = 2):

$$z\{\mathbf{s}_j; \delta(\mathbf{s}_j)\} = \min \left(\sum_{k=1}^h 1\{d_{jk} \leq \delta(\mathbf{s}_j)\}, 1 \right);$$

- d_{jk} : Distance between location \mathbf{s}_j and point source \mathbf{c}_k ;
- h : Total number of point sources.

Spatially-varying Radii and Exposure Effect Parameters

- Radii:

$$\Phi^{-1} \left(\frac{\delta(\mathbf{s}_j) - a}{b - a} \right) = \mathbf{w}(\mathbf{s}_j)^T \boldsymbol{\gamma} + \phi(\mathbf{s}_j);$$

- $\boldsymbol{\phi}^T = \{\phi(\mathbf{s}_1), \dots, \phi(\mathbf{s}_m)\} | \tau_\phi^2, \rho_\phi \sim \text{MVN}(\mathbf{0}_m, \tau_\phi^2 \Sigma(\rho_\phi)), \quad \Sigma_{ij}(\rho_\phi) = \exp\{-\rho_\phi \|\mathbf{s}_i - \mathbf{s}_j\|\};$
- Gaussian predictive process option also available for large m ;

- Effects:

$$\theta(\mathbf{s}_j) = \mathbf{q}(\mathbf{s}_j)^T \boldsymbol{\eta}.$$

Prior Information

$\beta_j \stackrel{\text{iid}}{\sim} N(\mathbf{0}, \sigma_\beta^2), j = 0, \dots, p_x - 1;$

- p_x : Length of \mathbf{x}_i vector (same for all i), including an intercept;
- Default setting: $\sigma_\beta^2 = 100^2$;

$\gamma_j | \tau_\phi^2 \stackrel{\text{iid}}{\sim} N\left(\mathbf{0}, \frac{1 - \tau_\phi^2}{p_w}\right), j = 0, \dots, p_w - 1;$

- p_w : Length of $\mathbf{w}(\mathbf{s}_j)$ vector (same for all j), including an intercept;

$\eta_j \sim N(0, \sigma_\eta^2), j = 0, \dots, p_q - 1;$

- p_q : Length of $\mathbf{q}(\mathbf{s}_j)$ vector (same for all j), including an intercept;
- Default setting: $\sigma_\eta^2 = 100^2$;

$\tau_\phi \sim \text{Uniform}(0, 1);$

$\rho_\phi \sim \text{Gamma}(a_{\rho_\phi}, b_{\rho_\phi});$

- Default setting: $a_{\rho_\phi} = b_{\rho_\phi} = 1$;

Gaussian specific: $\sigma_\epsilon^2 \sim \text{Inverse Gamma}(a_{\sigma_\epsilon^2}, b_{\sigma_\epsilon^2});$

- Default setting: $a_{\sigma_\epsilon^2} = b_{\sigma_\epsilon^2} = 0.01$;

Negative binomial specific: $r \sim \text{Discrete Uniform}[a_r, b_r];$

- Default setting: $a_r = 1, b_r = 100$.

Default Initial Values

- $\beta_j = 0$ for all j ;
- $\gamma_j = 0$ for all j ;
- $\eta_j = 0$ for all j ;
- $\tau_\phi = 0.50$;
- $\rho_\phi = -\ln(0.05) / \max\{||\mathbf{s}_i - \mathbf{s}_j||\}$;
- Gaussian specific: $\sigma_\epsilon^2 = \text{variance}(\mathbf{Y})$;
- Negative binomial specific: $r = 100$.

Additional Information

- `waic_info_indicator = 1`: Provides output needed to calculate Watanabe-Akaike information criterion;
- `fitted_info_indicator = 1`: Provides posterior samples of fitted values;
- `v`: Required input; an $\sum_{j=1}^m n_j$ -length vector with the location number that the observation is connected to (i.e., $1, \dots, m$);
- It is recommended to scale the matrix of spatial distances by the largest observed distance in the matrix prior to running the model.
- The `full_dists` matrix is a $m + k$ by $m + k$ matrix where the upper left m by m matrix describes the distances between the m observed locations, the bottom right k by k matrix describes the distances between the selected grid of $k \ll m$ locations (if the computational version is used), and the off-diagonal matrices describe the distances between the observed and grid locations (i.e., this full distances matrix is computed after stacking the observed locations and sampled grid locations into a single vector). If

the computational version is not needed, input the observed locations instead of the sampled locations when calculating these distances.